

ABSTRACT OF THE DISCLOSURE

A peer-to-peer storage system includes a storage coordinator that centrally manages distributed storage resources in accordance with system policies administered through a central administrative console. The storage resources are otherwise unused portions of storage media, e.g., hard disks, that are included in the devices such as personal computers, workstations, laptops, file servers, and so forth, that are connected to a corporate computer network. The devices are hereinafter referred to collectively as "storage nodes." The system thus centrally administers and manages the system-wide reuse of storage resources that are otherwise available only individually to the respective devices. The storage coordinator manages the distributed storage resources by assigning the nodes to various groups and allocating the storage resources on each of the nodes in a given group to maintaining dynamically replicated versions of the group files. The storage coordinator may also be dynamically replicated on other storage coordinators in a cluster, to prevent disruption of the system if one of the coordinators should fail. The files stored in the system are assigned to public and private storage domains, or "storage lockers," based in part on whether the files are shared access files or single-user access files. The storage coordinator manages the respective storage lockers as one or more "replication groups," which are essentially a mechanism for allocating storage resources to user files. The storage nodes perform file replication and synchronization operations by communicating directly, that is, peer-to-peer. The nodes use an XML message-based protocol that allows the nodes to perform the dynamic replication and synchronization operations as background tasks that are transparent to the user. The protocol also allows the nodes to handle lost or corrupted messages, the interruption of the replication and synchronization operations. The storage coordinator also manages distributed searches of file content on the network. In response to a query from a user, the storage coordinator selects one node from each storage group to search through the associated replication group files. The selected nodes perform their searches and report the search results back to the storage coordinator, which organizes the results and provides them to the user. Thereafter, in response to a request for various files by the user, the storage coordinator instructs the storage nodes that are near neighbors of the user to provide the requested

files. By careful selection of the nodes involved in a given distributed search operation, the storage coordinator minimizes the amount of network bandwidth consumed by the search operation.